

A Bottom-Up Scheme for User-Defined Feature Comparison in Ensemble Data

Richen Liu¹

Hanqi Guo²

Xiaoru Yuan^{1,3}

¹ Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University

² Mathematics and Computer Science Division, Argonne National Laboratory

³ Beijing Engineering Technology Research Center of Virtual Simulation and Visualization, Peking University

Abstract

Most of the existing approaches to visualizing the vector field ensembles are achieved by visualizing the uncertainty of individual variables of different simulation runs, for example, geometry distance, statistics, variability etc. However, the comparison of the derived feature or user-defined feature (e.g. vortex) is also of vital significance since they often make more sense according to the domain knowledge. In this paper, we present a new framework to extract user-defined feature from different simulation runs. Specially, we use a bottom-up searching scheme to help extract vortex with a user-defined shape. We further compute the geometry information including the size, and the geo-spatial location of the extracted vortex, and design some link views to compare them between different runs. Results show that our method is capable of extracting user-defined feature across different runs and comparing them spatially and temporally.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation;

Keywords: Ensemble Visualization, comparative visualization, user-defined feature

1 Introduction

Ensemble data are generated by different runs simulated by varied models or under different parameter constraints on the ensemble simulations. The analysis of ensemble data is of great importance to simulation science applications such as climate change analysis, operational weather forecast, computational fluid dynamics etc. Most of the existing approaches to analyzing the vector field ensembles mainly focus on the analysis of the uncertainty in different simulation runs. The uncertainty often can be quantified by the statistics of the variables, e.g. the expectation, the variance, the quartile, the confidence interval etc. All of these variables can be directly computed from the scalar field or the vector field of the ensembles. Nevertheless, the domain scientists often expect to see the derived feature instead of the individual variables in many scenarios. For example, the vortex or the air mass in climate study, which makes more sense than the individual variables, because the derived feature directly involve with the characteristic of the simulation. Specifically, we are interested in vortex in this paper, since it is a significant feature of climate research. The climate scientists require to study the location, size and duration time of the vortex between different simulation runs to analyze the atmospheric circulation. The changes of the atmospheric circulation are important indicators of climate change and are likely to have profound influences on ecosystems and societies [Reichler 2009]. Winds associ-

ated with the atmospheric circulation lead to transports of heat and moisture from remote areas and thereby modify the local characteristics of climate in important ways. Therefore, vortex can better reveal the characteristic of the ensemble simulations compared to the individual variables.

The intuitive explorations can largely help the scientists define the target vortex they are interested in. Sketching is an intuitive and convenient way to enable the scientists to define the vortex what they want. However, sketching is often restricted in a 2D space. It works well when the dimension of the data is only 2D. When it comes to a 3D case, one alternative method is to project the 3D curves into 2D and match them in 2D space. However, the accuracy will be decreased significantly due to the projection. In this paper, we introduce a bottom-up approach to search vortex in ensemble fields, which use sketch (2D) to filter vortexes, and select one vortex (3D) as template to find all the potential vortexes.

There are two main contributions of this paper. Firstly, we propose a framework to compare the user-defined derived feature in ensemble data. Compared to the traditional uncertainty visualization targeting to individual variables, the derived feature makes more sense according to the domain knowledge. Because the presence of a vortex in a simulation run at a given time-step is an important indicator of climate change and are likely to have profound influences on ecosystems and societies. Secondly, we take vortex as an example, and introduce a bottom-up approach to search vortex in ensemble vector fields.

2 Related Work

In this section, we survey the related works on uncertainty visualization in ensemble data and vortex region detection, respectively.

Visualization of vector field ensembles is one of key steps for simulation science applications such as climate change analysis, operational weather forecast, computational fluid dynamics, etc. The visualization of ensemble pathlines advected in the ensemble vector fields can reveal the characteristic feature of the simulation. It can help to improve the simulation model or achieve an optimized parameter configuration or the boundary conditions of the simulation. Obermaier et al [Obermaier and Joy 2014] have reviewed the few existing literatures, they have summarized some challenges on ensemble data visualization. Guo et al. [Guo et al. 2013] propose a parallel computation framework to compute the uncertainty between ensemble pathlines. The uncertainty is measured by the point-wise distances across different simulation runs.

Vortex region detection is based on scalar vortex region quantities that are used to define a vortex as a spatial region where the quantity exhibits a certain value range [Sahner 2009]. The classical vortex region detection method including Q -criterion [Hunt 1987] and λ_2 -criterion [Jeong and Hussain 1995]. The methods are deduced from the Navier Stokes equations that for a local pressure minimum. Both of the quantities are of limited applicability in some settings [Sahner 2009].

Recently, there are some line-based methods to effectively detect the vortex lines in vector field. A sketch-based approach [Wei et al. 2010] allows the user to sketch a 2D curve as template. A 3D field

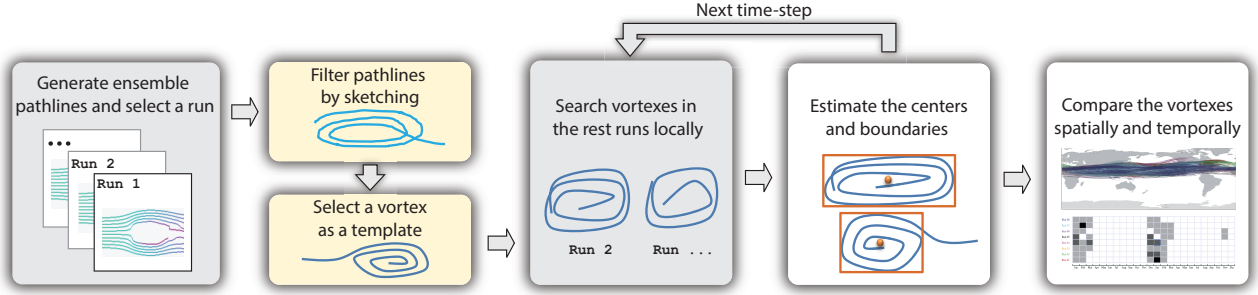


Figure 1: Pipeline of our comparative visualization. (1) Parallel pathline generation. (2) Select a base run, filter the pathlines to get template vortex by sketching (2D curve). (3) Select a vortex (3D curve) as the template and match in all the rest runs locally. (4) Estimate centers and boundaries of the found vortices and compute their occurrence time. (5) Visualization and comparison.

line whose view-dependent 2D projection is most similar to the template curve will be identified. Besides, a novel approach [Lu et al. 2013] was proposed to use distribution signature to extract the streamline with specific curvature and torsion. In this method, the streamline are firstly segmented, and their curvature distribution are then computed according to the segmentations, which will be summarized as a signature. The signature can reveal the geometry feature of the streamline and the distance between the signature are defined to support the subsequent streamline extraction and query.

3 Our Method

In this section, we will describe the primary two steps of our approach. The first one is vortex extraction, and the second one is vortex visualization and comparison.

Initially, we need to choose a vortex detection method to extract the vortex in the ensemble vector fields. Many traditional vortex region extraction methods (e.g. Q -criterion [Hunt 1987] and λ_2 -criterion [Jeong and Hussain 1995]) either focus on the pressure field or have limited applicability, as described in Section 2. The line-based methods [Wei et al. 2010] [Lu et al. 2013] do not need the additional scalar field and support user-defined sketching. Thus we choose line-based vortex detection method to compute the distance between pathlines.

Figure 1 shows the pipeline of our method. We use a bottom-up exploration scheme to search vortex in ensemble vector field. Firstly, we trace the pathlines from all simulation runs in parallel. We employ a slightly modified Map-Reduce-like parallel computing framework, named Dstep [Kendall et al. 2011] to trace the pathlines. Then select one run as base run. Secondly, we use a sketch-based method to extract vortex in the base run. It need the user to draw a 2D sketching curve to detect the potential vortex, and then replace the sketching curve with a vortex selected by the user, which becomes a new matching template. The user can use mouse to pick up the vortex end-points by ray intersection. Thirdly, the new template is utilized to compute its distance to all the pathlines from the rest runs. As a result, the vortices in all the runs can be identified by distance filtering. For different time steps, it need to conduct the matching process iteratively. Finally, the vortices at similar locations will be compared in different linked views, which are called peer vortices in this paper. The peer vortices are the vortices at similar geo-spatial locations, which should be extracted from different simulation runs. It is worth mentioning that if the user select another vortex in the base run as template, it needs to search its peer vortices following the pipeline again. In order to prevent searching the same vortex repeatedly, we use Google Protobuf library to serialize all the vortices onto disk, which can be loaded again for the subsequent visualization and comparison.

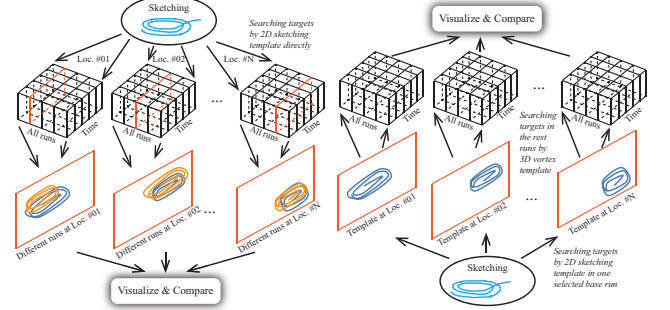


Figure 2: The comparison of exploration process between top-down method and bottom-up method.

3.1 Bottom-Up Searching Scheme

The goal of our method is to enable the user to define their target feature (e.g. vortex). Sketching is an intuitive and convenient way to achieve this goal. The users can define the size and the shape of the vortex in real time. The existing sketching method [Wei et al. 2010] to extract vortices from flow field is to sketch a curve in a 2D view, and take it as a template. Then make all the field lines match with the 2D template. 3D field line whose view-dependent 2D projection is most similar to the user drawing will be extracted. This exploration process is a classic top-down analysis for ensemble analysis, which is also called overview to detail analysis, as shown in the left part of Figure 2. Specifically, the overview will show all features across all runs extracted in the ensemble data. The detail view will provide the user comparison and analysis for different runs. The bottom-up approach is from detail to overview. We firstly extract vortex in a base run and make it as a template to extract the peer vortices. However, sketching is often restricted in 2D space. When it comes to 3D space, one alternative method is to project the 3D curves into 2D and match them in 2D space [Wei et al. 2010]. However, the accuracy will be decreased significantly. The curvature and angle of turn in 3D space and in a projected 2D space are quite different. In our bottom-up approach, this restriction can be avoided. We use sketch (2D) to search vortices just in the selected base run for each time of exploration, as shown in the right part of Figure 2. The result vortex (3D) in the base run will be considered as a new template to search the other vortices from the rest runs.

Furthermore, it is no sense to compare the vortices at different geo-spatial locations. For example, the vortex located in the North Pole in one simulation run should not be compared with that located in the South Pole in another run. The peer vortices are classified into a group with the same group label. In order to constrain the locations between peer vortices, it need to specify a spatial window to filter out the vortices which are out of the window. In our bottom-up method, it is easy to define the spatial window, since we can easily define the spatial window according to the location of the template vortex. In our experiment, we set the window size to be 1.5 times of the template vortex size. The scale is a parameter the user

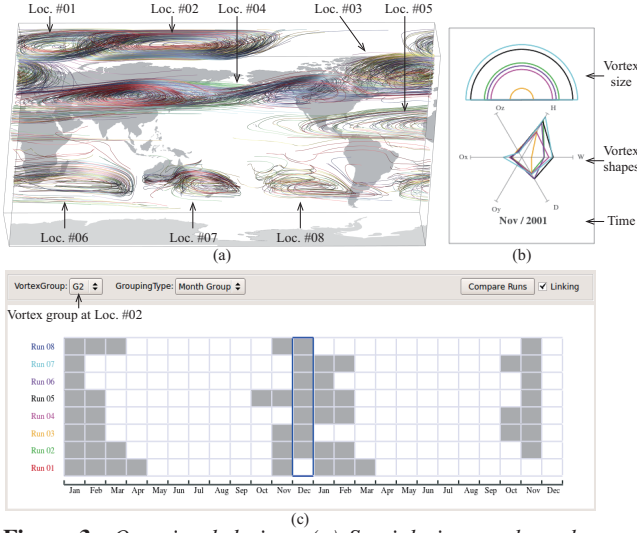


Figure 3: Our visual design. (a) Spatial view to show the geo-spatial information with the same group label. (b) A glyph-based card of the small multiples view. (c) Matrix view.

can adjust according to the searching results. In the matching step, we use the curvature and the angle of turn of each pathline to define the distance. As for each pathline, we firstly compute the histogram of its curvature and the histogram of the angle of turn along the time steps. Then we use a distance definition similar to the distribution-based streamline distance [Lu et al. 2013] to compute the distance. After that, the earth movers distance (EMD) [Rubner et al. 1998] is used to compute the histogram distance between two compared pathlines.

In a conclusion, there are three benefits for our bottom-up scheme compared to the top-down scheme. Firstly, we can avoid 2D projection by selecting a 3D vortex as the actual template. Secondly, it is more convenient to detect vortex at similar places by a spatial window. Thirdly, we can use different 3D template vortices at different locations to match the pathlines from the rest runs, because the vortex at different locations have different behavior.

3.2 Vortex Visualization and Comparison

The bottom-up vortex searching scheme allows to extract a series of groups of vortices. Different groups of vortices have different geo-spatial locations. In each group, the peer vortices have near geo-spatial locations. For the visualization part, we need to compare the extracted peer vortices spatially and temporally. The domain experts want to know the geo-location distribution, compare vortex location and vortex shapes across runs. Thus, we design three linked view to explore and compare the extracted vortices. They are spatial view, matrix view and small multiples view.

In the spatial view, each group of vortices (peer vortices in a group) are rendered to show their geo-spatial locations. In order to reduce visual clutter, it enables to select different group labels (as shown in the matrix view). The selected group of vortices will be drawn and other groups will be concealed when switching. User can also select a label named “All” to show all groups. Figure 3a shows 8 groups of vortices at the first time-step. We use color to encode different runs. In small multiples view, the estimated center and boundary of each vortex are visualized. The vortex size is encoded as the radius of each semi-circle on the top of Figure 3b, while the vortex shape, including width (W), height (H), depth (D), center location (Ox, Oy, Oz) are visualized in the star plot on the middle of Figure 3b. The matrix view is a pivot view because most of the interactive explorations are provided in this view. The x-axis is a time-line and the y-axis is all simulation runs. As for the current location group, the matrix cell is drawn in gray if it exists a vortex in the given run

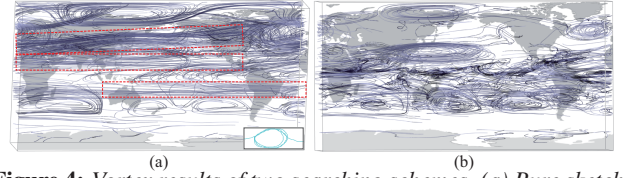


Figure 4: Vortex results of two searching schemes. (a) Pure sketch-based template searching (top-down scheme). (b) Vortex template searching (bottom-up scheme).

at the corresponding time, and drawn in white if there is no vortex, as shown in Figure 3c. It also supports to select different grouping types (month grouping or run grouping). The month grouping is to select one month for all runs, and the run grouping is to select one run for all time steps. The spatial view and the small multiples view will be changed accordingly if the user clicks on different matrix cells or different groups of cells. The button “Compare Runs” will trigger to show or hide the small multiples view.

4 Results and Cases

We conduct our experiments on a parallel environment. The platform is a PC cluster which consists of 8 nodes. Each node is equipped with two Intel Xeon E5520 CPUs which operate at 2.26 GHz and with 48 GB main memory. In our experiment, we employ a GEOS5 (Goddard Earth Observation System, Version 5) dataset simulated by this global atmospheric model. Its spatial resolution is $1^\circ \times 1.25^\circ$ for latitude-longitude grid coordinates, and 72 pressure levels (about 80 km) in the vertical direction. We employed the monthly average data with 8 runs from January, 2000 to December, 2001. Each run of the data is saved in 24 individual files corresponding to different time steps. The total size of this dataset is about 76 GB. To generate the pathlines, we use a Map-Reduce-like parallel computing framework, named Dstep [Kendall et al. 2011] to trace all the pathlines in the ensemble vector fields.

For the sake of making a comparison between the top-down scheme and the bottom-up scheme, we use sketch-based template to search vortex, the searching result of top-down scheme is shown in Figure 4a. The bottom-right of Figure 4a is a snapshot of the sketching template. It is easy to find that there are many straight lines highlighted in the red rectangles. The accuracy of pure sketch-based searching is relatively low since it need to project the 3D pathlines into 2D. While our bottom-up scheme use 3D vortex filtered by the sketching as new matching template, its accuracy is improved apparently. We can see that more vortices are found in Figure 4b. Besides, little straight lines are involved in this searching results.

Figure 5 shows the comparative visualization of our method. As described above, the user allows to choose three types of grouping method to specify the compared targets. For example, the user can select one run (Run 07) to show vortex in all time steps, as shown in Figure 5a. In the spatial view, we can see the geo-spatial distribution and the changes of vortex shapes along all time steps. In the small multiples view, we use focus and context technique to highlight the current run (Run 07), all other runs become the context and are drawn in gray. It is easy to find in this view that the size of vortex is increased slowly from Jan. 2000 to Mar. 2000, while decreased significantly from Jan. 2001 to Mar. 2001. This finding can help the user study the evolution of vortex for the selected run. In the star plot, it is convenient to find whether the vortex location and vortex size are outlier. The vortex depth (D) in Feb. 2001 is the most likely outlier in the star plot in Figure 5a, because it differs from other runs (in gray) to a large extent.

The user can also select one month (Mar. 2000) to compare all simulation runs, as shown in Figure 5b. In the spatial view, the 3D geo-spatial distribution of the vortices across all runs in Mar.



Figure 5: Results of the comparative visualization. (a) Grouping selection to select one run among all months. (b) Grouping selection to select one month across all runs. (c) Single cell selection to select a run with given time. (d) Matrix view of different locations.

2000 are similar. This finding can be also identified in the star plot since the variance of all 6 variables are relatively small. In the small multiples view, the card with the selected time will be highlighted in red rectangle. If we compare the runs for each month in the small multiples view, we can find the width (W) and O_x are the most possible to be outlier. This is a significant finding to control the simulation parameter. Furthermore, the user can also select a single cell in the matrix view to show the vortex from one specific run at a given time, as shown in Figure 4c.

Lastly, it is easy to find that the vortexes from almost all the location groups appear periodically in these two years, as shown in the matrix views in Figure 5(a-c) and in Figure 5d, this finding just follows the domain knowledge. However, we find the periodical information in location group No. 5 (Loc. # 05 in Figure 5d) is not that strong compared to other groups, this finding can be used to improve the simulation model and optimize the corresponding simulation parameters.

5 Discussion, Conclusions and Future Work

In this paper, we propose a new framework to visualize and compare the user-defined feature from vector field ensembles. This type of feature is derived from the original vector field. Compared to the existing methods, our user-defined feature can better describe the characteristics of the simulation according to the domain knowledge. Furthermore, we take vortex as an example and propose a bottom-up searching scheme to find vortex in different simulation runs. Although we just focus on vortex since it is quite significant for the climate research, our approach is not limited to the vortex. It supports any user-defined feature with different shapes for ensemble pathlines, because the proposed bottom-up scheme is a general approach to explore feature in ensemble data.

Nevertheless, our method has some limitations. If there are a large number of simulation runs, we need to make a better design to reduce visual clutter in spatial view, and need visual summary techniques to bundle simulation runs with similar behavior. Besides, although our scheme is more accurate, there are still some noise in the searching results, as shown in Figure 4b. It requires the user to interactively remove the noise pathlines. In the future, we plan to improve the interface to provide more information, e.g. the cell of the matrix view can be encoded with more information. Furthermore, the interface of small multiples should be improved.

Acknowledgements

This work is supported by NSFC No. 61170204 and the Strategic Priority Research Program - Climate Change: Carbon Budget and Relevant Issues of the Chinese Academy of Sciences Grant No. XDA05040205. This material is also based upon work supported by the U.S. Department of Energy, Office of Science, under contract number DE-AC02-06CH11357.

References

- GUO, H., YUAN, X., HUANG, J., AND ZHU, X. 2013. Coupled ensemble flow line advection and analysis. *IEEE Transactions on Visualization and Computer Graphics* 19, 12, 2733–2742.
- HUNT, J. C. R. 1987. Vorticity and vortex dynamics in complex turbulent flows. In *Canadian Society for Mechanical Engineering, Transactions (ISSN 0315-8977)*, vol. 11, 21–35.
- JEONG, J., AND HUSSAIN, F. 1995. On the identification of a vortex. *Journal of Fluid Mechanics* 285, 69–94.
- KENDALL, W., WANG, J., ALLEN, M., PETERKA, T., HUANG, J., AND ERICKSON, D. 2011. Simplified parallel domain traversal. In *Proceedings of the ACM Conference on Supercomputing*, 1–10.
- LU, K., CHAUDHURI, A., LEE, T., WEI SHEN, H., AND WONG, P. C. 2013. Exploring vector fields with distribution-based streamline analysis. In *Proceedings of IEEE Pacific Visualization Symposium*, 257–264.
- OBERMAIER, H., AND JOY, K. I. 2014. Future challenges for ensemble visualization. *IEEE Computer Graphics and Applications* 34, 3, 8 – 11.
- REICHLER, T. 2009. *Changes in the Atmospheric Circulation as Indicator of Climate Change*. Elsevier, ch. 7, 145–164.
- RUBNER, Y., TOMASI, C., AND GUIBAS, L. J. 1998. A metric for distributions with applications to image databases. In *The 6th International Conference on Computer Vision (ICCV)*.
- SAHNER, J. 2009. *Extraction of Vortex Structures in 3D Flow Fields*. PhD thesis, Magdeburg University.
- WEI, J., WANG, C., YU, H., AND MA, K. 2010. A sketch-based interface for classifying and visualizing vector fields. In *IEEE Pacific Visualization Symposium PacificVis*, 129–136.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, under contract number DE-AC02-06CH11357.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.